Virology 273, 341–350 (2000)
doi:10.1006/viro.2000.0434, available online at http://www.idealibrary.com on IDEAL®

Sequence Parameters That Determine Specificity of Binding of the Replication-Associated Protein to Its Cognate Site in Two Strains of *Tomato Leaf Curl Virus–New Delhi*

Anju Chatterji,* Udayan Chatterji,† Roger N. Beachy,* and Claude M. Fauquet*,[1]

*ILTAB/Donald Danforth Plant Sciences Center, University of Missouri St. Louis, CME, M307, 8001 Natural Bridge Road, St. Louis, Missouri 63121-4499; and †Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037

Received January 24, 2000; returned to author for revision April 25, 2000; accepted May 23, 2000

The DNA binding sites for the replication-associated protein (Rep) of two strains of tomato leaf curl virus from New Delhi (ToLCV-Nde) were identified using electrophoretic mobility shift assays (EMSAs). The Rep proteins of the two strains were found to exhibit sequence specificity in recognition of their cognate repeat motifs (iterons) in the origin, despite the fact that they share 91% sequence identity. Using a series of synthetic oligonucleotides as probes in EMSAs, the interaction of Rep protein with its binding site was found to be dependent on number, size, and sequence of the two iterons. Mutations in the sequence of the repeat motifs or alteration in the arrangement of the motifs compromised the ability of Rep protein to bind the DNA sequence and reduced accumulation of viral DNA in protoplasts, suggesting that binding of Rep protein to its cognate iterons is an essential step in virus replication. In addition, a difference in sequence of two base pairs in the binding site of two ToLCV-Nde strains was found to affect DNA binding by the corresponding Rep protein and replication of the virus DNA in protoplasts.

INTRODUCTION

Geminiviruses are plant DNA viruses characterized by twinned icosahedral particle morphology and covalently closed circular, single-stranded DNA as their genome (Stanley, 1991; Lazarowitz, 1992). Replication of the viral DNA occurs in the nuclei of infected cells via double-stranded DNA intermediates (Saunders *et al.,* 1991; Stenger *et al.,* 1991). The bipartite geminivirus genome contains open reading frames (ORFs) that are arranged in two divergent clusters separated by an intergenic region (IR). The IR contains sequences that are conserved between the two DNA components and are referred to as the common region (CR). The CR contains a GC-rich inverted repeat that is conserved in all geminiviruses and has the potential to form a stem-loop structure. These inverted repeats flank an AT-rich sequence of 11–16 bases that contains the conserved nonamer motif TAATATTAC.

The only geminivirus-encoded protein essential for viral DNA replication is the replication-associated protein (Rep). This protein is encoded by the ORF *AC1* and initiates rolling circle replication by a site-specific cleavage within the loop of the conserved stem-loop structure (Laufs *et al.,* 1995a,b). Rep is a multifunctional protein and is involved in both viral replication and transcriptional regulation (Sunter *et al.,* 1993; Fontes *et al.,* 1994a; Eagle *et al.,* 1994, 1997). The N-terminal region of the Rep protein has been shown to be important for DNA recognition and binding (Choi and Stenger, 1995, 1996; Jupin *et al.,* 1995) and in cleavage and ligation of the viral origin of replication. The C-terminal region of the Rep protein has a nucleoside triphosphate-binding domain (Laufs *et al.,* 1995b). Recently, discrete functional domains that are responsible for protein binding, cleavage, and oligomerization have been identified in the N-terminus of Rep protein of *Tomato golden mosaic virus* (TGMV) (Orozco *et al.,* 1997, 1998; Gladfelter *et al.,* 1997).

Rep interacts with at least two different DNA elements in the geminivirus origin of replication, a conserved nonanucleotide sequence containing a specific nick site (Heyraud-Nitschke *et al.,* 1995; Laufs *et al.,* 1995a; Stanley, 1995) and a directly repeated sequence motif located between the TATA box in the promoter of the *AC1* gene and the transcription start site (Fontes *et al.,* 1992, 1994a,b; Choi and Stenger, 1996). The Rep proteins encoded by different geminiviruses show specificity for the replication of their cognate genomes (Lazarowitz *et al.,* 1992; Fontes *et al.,* 1994b; Jupin *et al.,* 1995; Choi and Stenger, 1996). This specificity of origin recognition is determined, in part, by the high-affinity binding site of the Rep (Choi and Stenger, 1995, 1996) and the N-terminal domain of the Rep protein. In the case of *Tomato yellow leaf curl virus* (TYLCV), the N-terminal domain comprises the first 116 amino acid residues of the Rep protein (Jupin *et al.,* 1995).

Based on a phylogenetic and structural analysis of the IR from 30 different dicot-infecting geminiviruses, Ar-

[1] To whom correspondence and reprint requests should be addressed. Fax: (314) 516-4582. E-mail: iltab@danforthcenter.org.

0042-6822/00 $35.00

guello-Astorga *et al.* (1994) identified a series of sequence elements 8 to 12 nucleotides in length, which are repeated three to six times within the origin of replication. Their research indicated that the nucleotide sequence of the iterated elements (iterons) is generally virus specific and proposed that these iterons may represent specific sites for binding of Rep protein. The high-affinity binding sites for the Rep protein of TGMV, *Bean golden mosaic virus* (BGMV) (Fontes *et al.,* 1992, 1994a), and *Tomato leaf curl virus* (TLCV) (Behjatania *et al.,* 1998) have been mapped in the viral origin close to the TATA box and the conserved hairpin structure.

To determine the parameters that may influence specificity of binding of the Rep protein to its cognate site in the *ori,* we used two strains of *Tomato leaf curl virus* from New Delhi (ToLCV-Nde). These strains share 94% sequence identity but cause very different symptoms on tomato and *Nicotiana benthamiana* plants (Padidam *et al.,* 1995). Whereas the severe strain is characterized by severe puckering and downward leaf curling in the plants, the mild strain produces mild symptoms with minor leaf curl and no puckering of the leaves. In addition, the two strains do not support efficient replication of the heterologous DNA (less than 10% of the wild type levels) (Chatterji *et al.,* 1999), making them ideal experimental systems to investigate specificity of interaction between the Rep protein and its binding site in the origin of replication.

Previous studies (Chatterji *et al.,* 1999) indicated a possible interaction between amino acid 10 at the N-terminus of the Rep protein and the putative binding site in the origin that may determine specificity of replication between the severe and the mild strains of ToLCV-Nde. Exchange of the 10th amino acid between the Rep proteins of the two strains of ToLCV-Nde coupled with a change in the binding site sequence in the origin altered the replication of the two strains, suggesting that these components may influence the levels of viral replication and accumulation (Chatterji *et al.,* 1999).

In the current study our objective was to determine whether the repeat sequences identified earlier in the origin of replication of the two strains function as the binding sites for their respective Rep proteins and to define DNA sequence requirements for specificity of origin recognition of the two strains by using chimeric iteron sequences. Electrophoretic mobility shift assays (EMSAs) were performed using different synthetic oligonucleotides as probes or competitors to show specificity of binding in our assays. The nature and significance of DNA–protein interaction was studied *in vivo* using transient replication assays in tobacco protoplasts. We found that alteration with respect to sequence, size, or number of iterons that reduced binding by the Rep protein resulted in drastic reduction in virus replication.

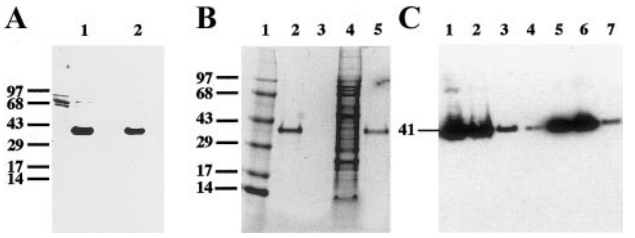

FIG. 1. Purification and analysis of 6xHis-tagged Rep proteins from two strains of ToLCV-Nde expressed in baculovirus AcNPV. (A) Western blot showing immunoprecipitation of Rep protein from crude lysates of Sf9 cells using anti-AC1 antiserum. The Rep proteins from the severe strain (A1, lane 1) and the mild strain (A2, lane 2) of ToLCV-Nde were expressed from the polyhedrin promoter of AcNPV in Sf9 cells and detected using anti-AC1 polyclonal antiserum. (B) Coomassie blue-stained gel showing the purified Rep proteins from the severe (lane 2) and mild (lane 5) strains of ToLCV-Nde. Lane 1 represents the marker and lane 4 shows the crude lysate from the pellet fraction. (C) Western blot using a polyclonal anti-AC1 antiserum. Stepwise eluates of the purified protein were collected from the affinity column in 20 mM Tris, 500 mM NaCl, and 500 mM imidazole (pH 7.9) and were detected using the anti-AC1 antiserum. Lanes 1–4 represent stepwise aliquots of the purified protein of the severe strain and the lanes 5–7 show similar fractions of the protein purified from the mild strain of ToLCV-Nde.

RESULTS

Expression and purification of the Rep protein from mild and severe strains of ToLCV

Previous studies (Chatterji *et al.,* 1999) showed that the severe (pMPA1) and the mild (pMPA2) strains of ToLCV-Nde exhibit specificity in replication of their cognate DNA. This selectivity is determined by interaction between amino acid residues at the N-terminus of Rep and its binding site in the intergenic region (IR). In this study, we examined the nature of this interaction directly *in vitro* by competitive DNA binding assays and *in vivo* by transient replication assays in tobacco protoplasts.

To isolate large amounts of Rep protein, the relevant sequences encoding the *AC1* gene of the mild and the severe strains of ToLCV-Nde were expressed from a polyhedrin promoter in a baculovirus expression vector. The recombinant baculovirus was used to infect Sf9 cells to obtain a high-titer virus stock and standardize optimal expression of the target protein. Cells were harvested at different time points after inoculation and protein was extracted after three cycles of freeze-thaw. The Rep protein from both the severe and the mild strain of ToLCV-Nde was detected in the crude insect cell lysates by immunoprecipitation using the polyclonal antiserum to AC1 (Fig. 1A, lanes 1 and 2).

The high-titer virus stock was used to infect High Five cells for large-scale purification of the target protein. The soluble protein extracts were loaded on a Ni-NTA column and the eluted fractions were analyzed by SDS–PAGE. The purified Rep protein had an estimated MW of 41 kDa in Coomassie blue-stained polyacrylamide gels (Fig. 1B, lanes 2 and 5) and its identity was further confirmed by

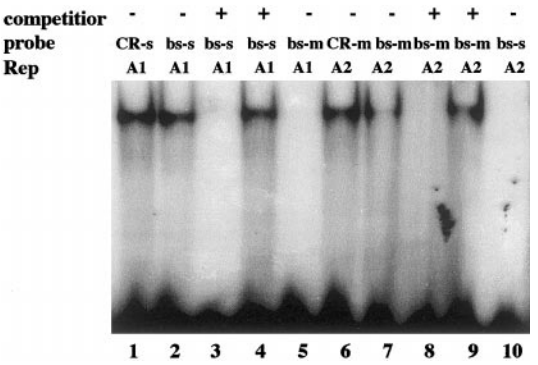

FIG. 2. Electrophoretic mobility shift assays showing the interaction of the Rep protein of severe (A1, lanes 1–5) and the mild (A2, lanes 6–10) strains of ToLCV-Nde with different common region fragments. CR-s and CR-m refer to the 52-bp common region fragment derived from the intergenic region of the viral DNA; bs-s and bs-m denote the 13-bp repeat motifs in the common region of the severe and the mild strain, respectively. ^{32}P-labeled DNA fragments (CR or bs) were incubated in the presence (+) or absence (−) of competitor DNA to test the specificity of binding. All reactions contained 200 ng of poly dI-dC and were analyzed on 4% polyacrylamide gels. The reactions in lanes 3 and 8 contained 50× molar excess of appropriate, unlabeled 13-bp DNA as the specific competitor and lanes 4 and 9 show the complex formation in the presence of 1000× molar excess of nonspecific competitor (pUC 18) DNA.

immunoblotting using AC1 polyclonal antiserum (Fig. 1C, lanes 1–7).

Rep proteins of the two strains bind a specific DNA sequence in the origin of replication and show selectivity in binding to their cognate iterons

Based on analogy with published reports describing binding sites of the Rep protein (Arguello-Astorga *et al.,* 1994; Fontes *et al.,* 1994a), putative repeat motifs close to the TATA box in the common regions of mild and severe strains of ToLCV were identified. The ability of the Rep proteins from the two strains to interact with their respective repeat motifs was determined in EMSAs. The repeat sequence in the severe strain was identified as GGTGTCTGGAGTC (nts 2640–2653), whereas in the mild strain the repeat motif identified was GGCGTCTGGCGTC (nts 2640–2653) (Chatterji *et al.,* 1999). We used the 13-bp sequence (nts 2640–2653) containing the repeat motifs as the probe in EMSAs with the purified Rep protein. As a control, the 52-bp fragment of the common region (nts 2614–2666) from the respective strains was used as a probe in similar assays.

A distinct DNA–protein complex was observed using both the common region (52-bp) fragment and the 13-bp repeat sequence of the severe strain (Fig. 2, lanes 1 and 2). These data suggest that the 13-bp sequence may represent the binding site for the Rep protein of the severe strain. Similarly, the Rep protein of the mild strain formed a complex with its respective 52-bp common region fragment as well as the 13-bp repeat motif sequence from the common region (Fig. 2, lanes 6 and 7).

The specificity of binding by the Rep proteins was tested in competition assays with radiolabeled probes and an excess of unlabeled DNA. Incubation of the Rep proteins with 50-fold molar excess of DNA of the same sequence abolished the formation of the complex completely (Fig. 2, lanes 3 and 8). On the other hand, a 1000-fold molar excess of a 13-bp sequence derived from the pUC 18 vector DNA did not affect the formation of the complex (Fig. 2, lanes 4 and 9). These results indicate that the interaction of Rep protein with its 13-bp sequence is highly specific. Since no obvious difference in terms of the mobility of the shifted complex was observed as a result of binding of the Rep protein to the 13-bp oligonucleotide or to the larger common region fragment containing the viral origin, the data suggested that the repeat sequences may constitute the high-affinity binding site for the respective Rep proteins of the two strains.

To determine whether the Rep protein from the mild strain could bind to the iteron sequence of the severe strain and vice versa, purified proteins from the two strains were incubated with labeled 13-bp sequence containing the heterologous repeat motifs in EMSAs. No DNA–protein complexes were observed when the severe strain Rep protein was incubated with the oligonucleotide containing the binding site sequence of the mild strain (Fig. 2, lane 5). Furthermore, the mild strain Rep protein did not bind to the 13-bp sequence comprising the repeat motifs of the severe strain (Fig. 2, lane 10). These results suggest that the Rep proteins exhibit high selectivity in binding to their cognate DNA sequences and two nucleotide differences in the sequence of the binding site can significantly affect the efficiency of binding.

Specificity of binding is related to the sequence, spacing, and number of iterons

Unlike the iteron sequences of the mild strain of ToLCV-Nde, the sequence of iterons that constitute the binding site of the severe strain are not identical. To better understand the basic DNA sequence requirements that contribute to or influence specificity of origin recognition, we made chimeric iteron sequences by exchanging individual motifs in the binding site of the two strains and determined the ability of Rep protein to bind them. In addition, the effect of each of these mutations on virus replication was determined in transient assays by assessing the ability of mutant DNA-A components to replicate in tobacco protoplasts.

Since the iterons constituting the binding site are repeated, they are referred to as 5′ or 3′, depending on their position in the common region. Mutant oligonucleotides were synthesized to determine whether the 5′ and the 3′ iterons contributed equally to the binding efficiency of the Rep protein and whether binding specificity

TABLE 1

Comparative Levels of *in Vitro* Binding and *in Vivo* Replication by the Origin Mutants

Mutant	Oligonucleotide sequence	Binding[a,b]	Replication[c]	
			ss[d]	sc[d]
CR-s	GGTGTCTGGAGTC	100	100	100
CR-m	GGCGTCTGGCGTC	100	100	89.5
IT-1/2(s)	GGTGTCTGGCGTC	100	100	93.4
IT-1/2(m)	GGTGTCTGGCGTC	<1[e]	<1	<1
IT-3/4(s)	GGGGTCTGGAGTC	13.4	28.6	<1
IT-3/4(m)	GGGGTCTGGAGTC	<1	<1	<1
IT-5/6(s)	GGGGTCTGGCGTC	<1	3.2	<1
IT-5/6(m)	GGGGTCTGGCGTC	<1	<1	<1
IT-7/8(s)	GGCGTCTGGGGTC	<1	2.8	<1
IT-7/8(m)	GGCGTCTGGGGTC	18.4	<1	<1
IT-9/10(s)	GGTGTCTGGTGTC	102.6	108	103.9
IT-11/12(s)	GGAGTCTGGAGTC	87.2	98.6	98.8
IT-13/14(s)	GGTGTCTTTTTTGGAGTC	<1	4.5	<1
IT-13/14(m)	GGCGTCTTTTTTGGCGTC	<1	<1	<1
IT-15/16(s)	GGTGTCGGAGTC	<1	<1	<1
IT-15/16(m)	GGCGTCGGCGTC	<1	<1	1.5
IT-17/18(s)	GGTGTC	29.4	24.6	<1
IT-17/18(m)	GGCGTC	8.5	17.6	<1
IT-19/20(s)	GGTGTCTGGTGTCTGGTGTCTGGTGTC	—[f]	75.9	47.1
IT-19/20(m)	GGCGTCTGGCGTCTGGCGTCTGGCGTC	—[f]	62.2	36.5
IT-21/22(s)	GGCGTCTGGTGTC	100	78.2	45.9
IT-21/22(m)	GGCGTCTGGTGTC	14.8	8.4	<1
IT-23/24(s)	GGTGTCTGGGGTC	100	76.9	45.3
IT-23/24(m)	GGTGTCTGGGGTC	<1	<1	<1

[a] The values shown represent the amount of radioactivity (%) bound with respect to the wild type.[b] (100%) in the shifted DNA–protein complex band as a result of the Rep protein binding to the ^{32}P labeled DNA in gel shift assays.

[b] The amount of radioactivity bound in the complex shifted as a result of Rep protein binding to its respective CR sequences was assigned a value of 100.

[c] The values shown are average (%) amounts of single stranded (ss) and supercoiled (sc) viral DNA detected in four independent protoplast transfections per mutant. Standard error values between different transfections were in the range of ±2–5%.

[d] The amounts of viral DNA observed in protoplasts inoculated with the wild type DNA-A of the severe or the mild strain were assigned a value of 100.

[e] Too low for accurate quantification because of background error.

[f] The radioactivity was distributed in several complexes and was therefore not directly comparable with other mutants.

between the strains could be altered by exchanging the appropriate iterons. These oligonucleotides were used as probes in EMSAs to test their ability to form a complex with the purified Rep proteins. The efficiency of binding by the Rep protein to these chimeric iteron sequences was detected by *in vitro* assays and their effect on replication ability of the viral DNA was assessed *in vivo* by quantifying the amount of radioactivity bound in EMSAs and Southern blots, respectively. These values are provided in Table 1.

Mutations affecting the sequence of iterons. Four sets of oligonucleotides containing different combinations of iterons were tested. All of these chimeric iterons represented either one or two base-pair substitutions at the 3rd or the 10th relative to the cognate 13-mer iteron sequence of both the severe and the mild strain of ToLCV-Nde.

The Rep protein of the severe strain formed a strong complex with the sequence, GG**T**GTCTGG**C**GTC (IT 1/2, Fig. 3A, lane 1) but a faster-migrating weak complex was also observed. However, the Rep protein of the mild strain did not bind to this probe (Fig. 3B, lane 1). In protoplasts, the severe strain background with this mutant origin was able to support replication of the virus to nearly wild type levels but the mild strain accumulated barely detectable levels of viral DNA (Fig. 3C, lanes 2 and 13).

The Rep protein of the severe strain formed two very weak complexes with the probe GG**G**GTCTGG**A**GTC (IT 3/4, Fig. 3A. lane 2), and the Rep protein from the mild strain failed to generate a complex with this probe *in vitro* (Fig. 3B, lane 2). In tobacco protoplasts very low levels of viral DNA accumulation were observed in the case of severe strain mutant (Fig. 3C, lane 3). The mild strain mutant did not replicate to detectable levels (Fig. 3C, lane 14).

Neither of the Rep proteins detectably bind to the sequence GG**G**GTCTGG**C**GTC (IT 5/6, Fig. 3A, lane 3 and

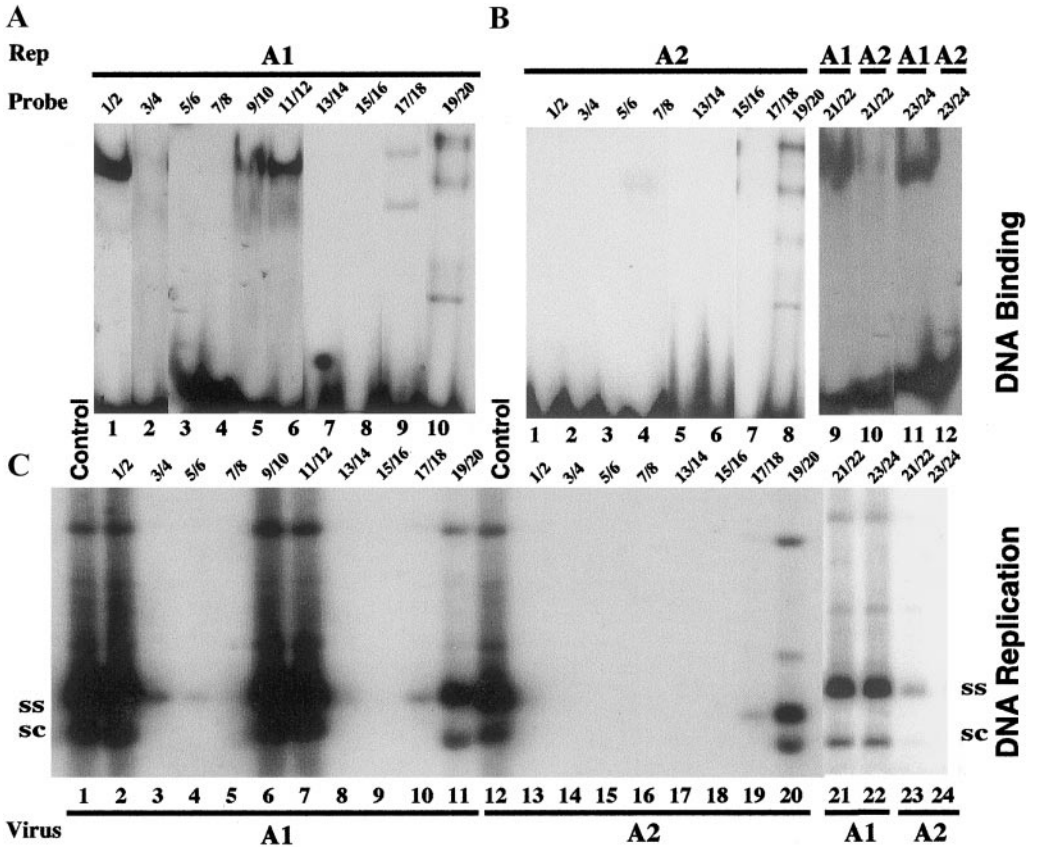

FIG. 3. DNA sequence requirements for binding by the Rep protein. (A) Labeled synthetic oligonucleotides with variations in the sequence and arrangement of iterons were used as probes to analyze their effect on binding by the Rep protein of severe strain of ToLCV-Nde. The key to the sequence of iterons used as probes is as follows: IT 1/2 (5′ severe, 3′ mild, lane 1); IT 3/4 (5′ different, 3′ severe, lane 2); IT 5/6 (5′ different, 3′ mild, lane 3); IT 7/8 (5′ mild, 3′ different, lane 4); IT 9/10 (5′ severe repeated, lane 5); IT 11/12 (3′ severe repeated, lane 6); IT 13/14 (spacing within the iterons is increased to 6 nucleotides, lane 7); IT 15/16 (no spacing between the iterons, lane 8); IT 17/18 (5′ monomer iteron, lane 9); IT 19/20 (5′ monomer repeated four times, lane 10). (B) Labeled synthetic oligonucleotides with variations in the spacing and number of iterons were used as probes to analyze their effect on binding by the Rep protein of the mild strain of ToLCV-Nde. The key to the sequence of iterons used as probes is as follows: IT 1/2 (5′ severe, 3′ mild, lane 1); IT 3/4 (5′ different, 3′ severe, lane 2); IT 5/6 (5′ different, 3′ mild, lane 3); IT 7/8 (5′ mild, 3′ different, lane 4); IT 13/14 (spacing within the iterons is increased to 6 nucleotides, lane 5); IT 15/16 (no spacing between the iterons, lane 6); IT 17/18 (5′ monomer iteron, lane 7); IT 19/20 (5′ monomer repeated four times, lane 8); IT 21/22 (5′ mild, 3′ severe, lanes 9 and 10); IT 23/24 (5′ severe, 3′ different, lanes 11 and 12). (C) Replication of Rep protein binding site mutants. Plasmids (2 μg) containing the viral replicons mutated at their binding site sequence in the origin were electroporated into tobacco protoplasts. Total DNA was isolated 48 h after transfection, resolved on agarose gels, and analyzed by Southern hybridization using ^{32}P-labeled AC1 DNA fragment (nts 2113 to 2695) as a probe. The single-stranded (ss) and the supercoiled (sc) forms of the viral DNA are indicated. The virus mutants were given identical names as the oligonucleotides used to alter their iteron sequence for the sake of convenience. The key to the mutants is as follows: Lanes 2 and 13: IT 1/2; lanes 3 and 14: IT 3/4; lanes 4 and 15: IT 5/6; lanes 5 and 16: IT 7/8; lane 6: IT 9/10; lane 7: IT 11/12; lanes 8 and 17: IT 13/14; lanes 9 and 18: IT 15/16; lanes 10 and 19: IT 17/18; lanes 11 and 20: IT 19/20; lanes 21 and 23: IT 21/22; lanes 22 and 24: IT 23/24. Lanes 1 and 12 represent the wild type controls for the severe and the mild strain of ToLCV-Nde, respectively.

Fig. 3B, lane 3) *in vitro*. Very low levels of single-stranded viral DNA accumulation were detected in tobacco protoplasts when inoculated with the severe strain mutant but the mild strain mutant failed to replicate (Fig. 3C, lanes 4 and 15).

The Rep protein of the severe strain did not bind to the sequence GG**C**GTCTGG**G**GTC (IT 7/8) *in vitro* (Fig. 3A, lane 4) but the Rep protein of the mild strain formed two closely migrating weak complexes with this probe (Fig. 3B, lane 4). In tobacco protoplasts, the severe strain mutant accumulated very low amounts of viral DNA (2.84%, Table 1), whereas in the case of the mild strain mutant no viral DNA accumulation was observed (Fig. 3C, lanes 5 and 16). The low level of viral DNA associated with the severe strain mutant IT 7/8 was not the result of contamination from an adjacent lane.

The iterons comprising the binding site of the severe strain Rep protein are not identical repeats. To determine whether either of the iteron sequences influence the binding efficiency, synthetic oligonucleotides were designed with GG**T**GTCTGG**T**GTC (IT 9/10) and GG**A**GTCTGG**A**GTC (IT 11/12) as perfect repeats and tested in EMSAs for their capacity to bind the Rep protein of the severe strain. In gel shifts, the Rep protein bound to GGTGTCTGGTGTC, as visualized by a retarded complex, but the binding to GGAGTCTGGAGTC was weaker than

the mutant IT 9/10 (Fig. 3A, lanes 5 and 6; Table 1). In protoplasts, both mutants replicated viral DNA similar to the wild type controls (Fig. 3C, lanes 6 and 7).

The Rep protein of the severe strain formed a strong complex with the probe GGCGTCTGGTGTC (IT 21/22), but in the case of the Rep protein of the mild strain, a weak complex could be distinguished (Fig. 3B, lanes 9 and 10, respectively). In protoplasts, the severe strain mutant replicated to high levels but the mild strain mutant accumulated low amounts of viral DNA (Fig. 3C, lanes 21 and 23).

The Rep protein of the severe strain bound to the probe GGTGTCTGGGGTC (IT 23/24) with high affinity (Fig. 3B, lane 11) and accumulated viral DNA in tobacco protoplasts (Fig. 3C, lane 22). However, the mild strain Rep protein did not bind to this probe *in vitro* (Fig. 3B, lane12) and did not replicate viral DNA in protoplasts (Fig. 3C, lane 24).

Mutations affecting the spacing within the iterons. The iterated motifs in both the severe (GGTGTCTGGAGTC) and the mild (GGCGTCTGGCGTC) strains of ToLCV-Nde are separated by a single nucleotide. To find out if the spacing between the iterons was significant for origin recognition by the Rep protein, the distance between the iterons was either increased to six nucleotides (IT 13/14) or reduced to none by deleting the single T nucleotide (IT 15/16) between the iterons. DNA–protein complexes were not observed when the spacing between the two repeats was either increased to six bases or decreased to zero for both severe and mild strains (Fig. 3A, lanes 7 and 8; Fig. 3B, lanes 5 and 6). In protoplasts, the mutant IT 13/14 of the severe strain accumulated very low levels of single-stranded DNA only (4.4%, Table 1), whereas the mutant IT 15/16 did not replicate at all (Fig. 3C, lanes 8 and 9). Again, the low level of viral DNA associated with the severe strain mutant IT 13/14 was not the result of contamination from an adjacent lane. Neither of the two mutants in case of the mild strain showed detectable levels of virus replication (Fig. 3C, lanes 17 and 18).

Mutations affecting the number of iterons. The Rep protein of the severe strain was unable to bind as efficiently (in comparison to the wild type bs-s sequence) if only a single iteron (IT 17/18, GGTGTC) was used as a probe (Fig. 3A, lane 9; Table 1). When the number of repeat motifs was doubled (IT19/20, GGTGTCTGGAGTCTGGTGTCTGGAGTC), multiple complexes with retarded mobility were observed (Fig. 3A, lane 10). The Rep protein of the mild strain did not bind as efficiently to a monomer GGCGTC (IT17/18) (Fig. 3B, lane 7). Doubling the number of repeat sequences (IT 19/20, GGCGTCTGGCGTCTGGCGTCTGGCGTC) resulted in the appearance of several shifted bands (Fig. 3B, lane 8). In transient assays, the mutants containing a single iteron did not replicate very well and only little accumulation of single-stranded DNA was observed (Fig. 3C, lanes 10 and 19). However, the mutants containing twice the number of repeat motifs in their origin were able to replicate (Fig. 3C, lanes 11 and 20), although not as efficiently as the wild type controls (Fig. 3C, lanes 1 and 12).

DISCUSSION

This research defines DNA sequences in the viral origin of replication that are specifically recognized by the Rep protein of two strains of ToLCV-Nde and demonstrates that binding of the Rep protein to their cognate sequences may constitute an important step in viral replication. In addition, the binding of the Rep proteins to their cognate iterons is found to be highly specific between the strains and is dependent on several criteria, including the sequence, spacing, and the number of iterons. Further, evidence is provided that any mutation in the iteron motifs that affects DNA binding *in vitro* impacts viral DNA accumulation *in vivo*.

Previous studies (Chatterji *et al.,* 1999) identified 13-mer repeat sequences in the common region of ToLCV-Nde that are involved in interaction with the Rep protein and are essential for virus replication. Competitive DNA binding assays using purified Rep proteins established that the Rep protein of ToLCV-Nde (severe strain) specifically binds to the iterated sequence, 5′ GGTGTCTGGAGTC located on the 5′ end of the TATA box between positions 2640 and 2653 and is conserved between the DNA-A and DNA-B components. Similar iterated sequences have been found in the common region of several geminiviruses (Arguello-Astorga *et al.,* 1994) and some have been verified in biochemical assays to act as Rep protein-specific binding sites (Fontes *et al.,* 1992, 1994a; Bejhatania *et al.,* 1998). The specificity of binding was confirmed by competition assays. A 50-fold molar excess of the homologous probe abolished the formation of a DNA–protein complex, yet a 1000-fold molar excess of heterologous DNA (pUC18) did not affect the binding of the Rep protein, indicating specificity of the complex formed. The binding efficiency of the Rep protein to the 13-bp oligonucleotide or to the 52-bp common region fragment was indistinguishable. These data suggested that the complex observed as a result of Rep protein binding to the 13-bp sequence is authentic and the two repeat sequences represent the binding site of the Rep protein in both the mild and severe strains of ToLCV-Nde.

The sequences of repeat motifs that constitute the binding site differ between the mild and the severe strain by two base pairs, GG**T**GTCTGG**A**GTC for the severe strain versus GG**C**GTCTGG**C**GTC for the mild strain. In EMSAs the Rep protein of the two strains did not form a complex with the 13-mer DNA sequence of the heterologous strain, suggesting a high degree of specificity between the Rep proteins for their cognate sites. This observation is not surprising since there are several examples among geminiviruses documenting high specificity in replication of cognate genomes. As well, pseu-

dorecombinants can be formed only between related viruses or strains of a specific virus (Lazarowitz *et al.,* 1992; Stanley *et al.,* 1990; Frischmuth *et al.,* 1993). Fontes *et al.* (1994b) reported that the Rep protein of BGMV can specifically recognize its cognate site but is unable to bind to the corresponding binding site of a related virus such as TGMV. Similarly, the inability of Worland, CFH, and the Logan strains of *Beet curly top virus* (BCTV) to transcomplement replication of each other might be related to specific recognition of the binding site by the corresponding Rep proteins (Choi and Stenger, 1995; Stenger, 1998). Together, these results highlight the importance of specific interaction between the Rep protein and its cognate site in the replication of geminiviruses. In addition, it is important to emphasize that precise recognition motifs of Rep proteins may be different between the viruses and are not interchangeable, thereby maintaining specificity between closely related viruses and strains of the same virus.

Direct evidence for the role of Rep protein binding to the repeat sequences in geminivirus replication was obtained by the mutation of these sequences in the tomato leaf curl virus origin and testing mutants for their ability to replicate in tobacco protoplasts. We found that the ability of the Rep protein to bind to its cognate site was correlated to the ability of the ToLCV-Nde DNA to replicate in transient assays. Of the 20 mutants tested in this study, 16 mutants that did not form a DNA–protein complex in the EMSAs did not replicate viral DNA. However, three of the 20 mutants studied did not show any detectable *in vitro* binding, yet accumulated very low levels of single-stranded viral DNA in protoplasts (2.8–4.4%). Given the barely detectable levels of replication in the case of these three mutants, it is possible that the Rep might have bound the mutated iteron sequences as well but it could not be detected because of sensitivity limitations of EMSAs. On the other hand, the mutant IT 7/8 (m) did bind the mutated iteron sequences *in vitro* (18.4%), yet did not accumulate detectable levels of viral DNA in transient assays, implying the involvement of additional factors in virus replication. Together, these data suggest that recognition or efficient binding to the iteron motifs by the Rep protein is required for optimal viral DNA replication and accumulation to occur.

The sequence requirements that govern the recognition specificity of the Rep protein for its cognate site were determined by altering sequence, spacing, and number of iterons. Mutagenesis of the Rep protein binding site showed that, although both repeat motifs are essential for optimal DNA binding to occur *in vitro,* the efficiency of binding was more dependent on the sequence of its cognate 5′ iteron in the case of the Rep protein from the severe strain. The Rep protein of the mild strain, however, did not bind efficiently unless both the 5′ and 3′ iterons were homologous, indicating that both the iterons might be essential to achieve maximal binding *in vitro.* This observation was reinforced by the behavior of mutants IT 17/18 (s and m) that contain monomeric copies of the 5′ iteron, yet do not bind DNA efficiently *in vitro* (Table 1). These results were also reflected by the reduced levels of viral DNA accumulation in protoplasts, suggesting that efficiency of replication may be correlated to the Rep protein binding to its specific site in the origin.

The repeat sequences comprising the binding site in the severe strain of ToLCV-Nde are not identical, which led us to examine whether the presence of identical repeats is important as the binding site or whether one iteron is more significant than the other in terms of the efficiency of binding. More important, can the exchange of their iteron sequences alter the specificity of origin recognition between the two Rep proteins? Our results indicate that the two iterons, even when they are identical, make different contributions to the efficiency of binding, as has been shown for TGMV (Fontes *et al.,* 1994a). In the case of the severe strain, the presence of a homologous 5′ iteron appears to be more important for binding (IT 1/2), since substitution by a heterologous 5′ iteron did not allow efficient complex formation (IT 3/4), even though the 3′ iteron was homologous. Although the Rep proteins of both strains could bind to a monomeric copy of their respective 5′ iteron, the binding was not efficient, as evidenced by the gel shift assays and low levels of single-stranded DNA replication in protoplasts.

It would appear that a 3′ repeat sequence or a 13-mer is required to promote efficient binding by the Rep protein. This observation was substantiated by the mutant IT 1/2 (s) that replicated and showed efficient binding by the A1 Rep protein, which did not bind to bs-m. In addition, the mutant IT 3/4 (s) did not form a strong complex in EMSAs and accumulated only 28.6% of the single-stranded DNA, suggesting yet again that the sequence of the 5′ iteron is more important than its 3′ counterpart. This observation is contrary to what has been observed in the case of TGMV (Orozco *et al.,* 1998), wherein the authors reported that the 3′ iteron contributes more to the efficiency of binding than does the 5′ iteron. The significance of this difference is unclear, considering the fact that TGMV and ToLCV-Nde originated in different parts of the world and are not closely related. In addition, since we did not test an exhaustive number of mutants with all possible combinations of iteron sequences, we cannot conclude unequivocally that the 5′ iteron is more important than the 3′ iteron. In general, the severe strain of ToLCV-Nde was found to be more tolerant than the mild strain to changes at the 3rd and the 10th nucleotide of the 13-mer repeat sequence. Overall, our results indicate that the efficiency of binding of the Rep protein to its cognate site can be correlated to the levels of viral DNA accumulation in protoplasts.

The effect of deletion of a single base pair between the two repeat sequences resulted in negligible viral DNA

replication levels, indicating that the ToLCV-Nde origin is sensitive to changes in the spacing of the iterons. Similar observations were also reported in the case of TGMV, in which the spacing within and between the repeats was found to be important for AC1 binding and origin function (Orozco *et al.,* 1998). Because the Rep protein might bind as a dimer (Fontes *et al.,* 1994a), it is possible that the proximity of the two repeat motifs is congenial for binding to occur and any alterations with respect to spacing between the two iterons does not allow efficient recognition. Doubling the number of iterons resulted in the appearance of multiple bands, which may indicate protein–protein interaction was facilitated as a result of multimerization of the Rep protein, without causing any drastic reduction to the accumulation of viral DNA.

These results imply that recognition of cognate iterons may represent an important step in the replication process, although there might be other interactions between the iteron sequences and, possibly, other yet-to-be-identified proteins that recognize or bind them, which might have some role to play in the replication process.

MATERIALS AND METHODS

Expression of Rep proteins of ToLCV-Nde

The full-length *AC1* gene from the severe and the mild strains of ToLCV-Nde was amplified from pMPA1 (DNA-A of the severe strain, ToLCV-Nde) and pMPA2 (DNA-A of the mild strain, ToLCV-Nde) (Padidam *et al.,* 1995). The amplified sequence was ligated between *Bam*HI and *Hin*dIII sites in baculovirus expression vector pBAC4x-1 (Novagen, Madison, WI), resulting in an in-frame fusion of the *AC1* gene sequence with the vector sequence encoding a methionine and six histidine residues under the *polh* promoter. The clones were identified and confirmed by restriction digestion and sequence analysis.

Recombinant baculovirus was isolated by cotransfecting 0.5 μg of recombinant plasmid with 1 μg of linearized *Autographa californica* nuclear polyhedrosis virus DNA (Smith and Summers, 1978) into *Spodoptera frugiperda* Sf9 cells (Summers and Smith, 1987). Recombinant viruses were plaque-purified and a high-titer stock was prepared. Large-scale purification of the target protein was done in *Trichoplusia ni* High Five cells (Invitrogen, Carlsbard, CA).

Purification of Rep proteins

High five cells were harvested 60 h postinfection by centrifugation at 3000 *g* for 10 min. The pellets were washed in 1× PBS and suspended in ice-cold 1× binding buffer (5 mM imidazole, 0.5 M NaCl, and 20 mM Tris, pH 7.9). The cells were lysed by three cycles of freeze-thaw and the lysate was clarified at 17,000 *g* for 30 min. The resulting supernatant was loaded on a Ni-NTA column (Novagen) previously equilibrated with binding buffer and washed with 10 column volumes of wash buffer (70 mM imidazole, 0.5 mM NaCl, and 20 mM Tris, pH 7.9). The protein was eluted with 1 M imidazole, 0.5 mM NaCl and 20 mM Tris, pH 7.9. The eluted fractions were dialyzed against 20 mM Tris, pH 7.9, 150 mM NaCl to remove imidazole, and concentrated using Centricon filters (Amicon, Beverly, MA) and protein concentration was estimated using Bradford's reagent (Bio-Rad, Hercules, CA).

Electrophoretic mobility shift assays

ToLCV-Nde-specific primers were used to amplify a 52-bp fragment from the IR of the virus genome. This fragment contains the iterons, the transcription start site, as well as the TATA box and the conserved hairpin sequence. The amplified fragment was end-labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase, purified on polyacrylamide gels and was used as a probe in the EMSAs. The 13-mer oligonucleotides containing the potential binding sites (underlined) for the Rep proteins of the two strains were synthesized and annealed to their complementary strands. These two oligonucleotide probes were named bs-m (5′-GGCGTCTGGCGTC-3′) for the mild strain and bs-s (5′-GGTGTCTGGAGTC-3′) for the severe strain. The final concentrations of the probes were 500 pM (30,000 cpm). The concentration of competitor DNA used was 100 pM per reaction. Probe and competitor DNAs were purified on a Sephadex G-25 column and quantified by a scintillation counting, followed by dilution to 30,000 cpm for the binding assays. The sequences of the synthetic oligonucleotides used as probes or competitors in the EMSAs are listed in Table 1.

The binding assays (EMSAs) were performed using the purified Rep protein from the two strains. The binding reactions contained 500 ng of pure protein, 1 ng of labeled DNA, and 0.2 μg of poly dI-dC. Binding buffer contained 20 mM HEPES, pH 7.5, 60 mM KCl, 1 mM DTT, and 15% glycerol. Reactions were incubated at 25°C for 30 min and the complexes were resolved on 4% polyacrylamide gels in 0.25× TBE. The gels were dried on Whatman paper and autoradiographed. Comparative efficiency of binding between different mutants was analyzed by quantifying the amount of radioactivity in the retarded bands using the phosphorimager (Molecular Dynamics, Sunnyvale, CA).

Immunoprecipitation assays

Rep protein was immunoprecipitated from 50 μg of Sf9 cell extract using 10 μg of polyclonal anti-AC1 antiserum and rabbit anti-mouse IgG coupled to Sepharose beads. The immunocomplexes were resolved by SDS–PAGE and detected by immunoblotting (Towbin *et al.,* 1979) using a rabbit polyclonal anti-AC1 antiserum and an anti-rabbit goat antibody conjugated with horseradish peroxidase. The peroxidase activity was detected using

chemiluminescent Supersignal substrate (Pierce, Rockford, IL).

Construction of mutants

A 146-bp common region fragment from *Nco*I to *Ssp*I sites from both severe and mild strains of ToLCV-Nde was amplifed and cloned in pBS (SK-) vector. Site-directed mutations were made in the iterated sequences using overlapping PCR (Horton, 1994). Following confirmation of mutations by sequencing, the common region fragments were recloned into the DNA-A of the respective strains. For convenience, the mutants were given names identical to the synthetic oligonucleotides used to create the nucleotide changes in the iteron sequence.

Transient replication assays

Protoplasts isolated from *Nicotiana tabacum* BY2 cells were electroporated and cultured according to published methods (Watanabe, 1987). Transfections were done using 2 μg each of the wild type or mutant replicon DNA containing a partial tandem copy of pMPA1, pMPA2, or their derivatives (Chatterji *et al.,* 1999). The mutant replicon contained sequence alterations in the iteron sequences in the viral origin. Total DNA from the protoplasts was extracted 48 h after transfection (Dellaporta *et al.,* 1983; Mettler *et al.,* 1987) and analyzed for viral DNA accumulation by Southern blotting (Chatterji *et al.,* 1999). Comparative levels of viral replication were estimated by quantification using a phosphorimager (Molecular Dynamics).

ACKNOWLEDGMENTS

A.C. expresses her sincere gratitude to Dr. John H. Elder for providing guidance and unlimited resources to do the baculovirus expression and protein purification work in his laboratory. This work was supported by financial assistance from MAHYCO, Jalna, India (Grant 5-98378) and IRD, Paris, France.

REFERENCES

Arguello-Astorga, G. R., Guevara-Gonzalez, R. G., Herrera-Estrella, L. R., and Rivera-Bustamante, R. F. (1994). Geminivirus replication origins have a group specific organization of iterative elements: A model for replication. *Virology* **203,** 90–100.

Behjatnia, S. A., Dry, I. B., and Rezaian, M. A. (1998). Identification of the replication associated protein binding domain within the intergenic region of tomato leaf curl geminivirus. *Nucleic Acids Res.* **26,** 925–931.

Chatterji, A., Padidam, M., Beachy, R. N., and Fauquet, C. M. (1999). Identification of replication specificity determinants in two strains of tomato leaf curl virus from New Delhi. *J. Virol* **73,** 5481–5489.

Choi, I. R., and Stenger, D. C. (1995). Strain-specific determinants of beet curly top geminivirus DNA replication. *Virology* **206,** 904–912.

Choi, I. R., and Stenger, D. C. (1996). The strain-specific *cis* acting element of beet curly top geminivirus DNA replication maps to the directly repeated motif of the *ori*. *Virology* **226,** 122–126.

Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983). A plant DNA minipreparation: Version II. *Plant Mol. Biol. Rep.* **1,** 19–21.

Eagle, P. A., and Hanley-Bowdoin, L. (1997). *Cis* elements that contribute to geminivirus transcriptional regulation and the efficiency of DNA replication. *J. Virol.* **71,** 6947–6955.

Eagle, P. A., Orozco, B. M., and Hanley-Bowdoin, L. (1994). A DNA sequence required for geminivirus replication also mediates transcriptional regulation. *Plant Cell* **6,** 1157–1170.

Fontes, E. P. B., Eagle, P. A. Sipe, P. S. Luckow, V. A., and Hanley-Bowdoin, L. (1994a). Interaction between a geminivirus replication protein and origin DNA is essential for viral replication. *J. Biol. Chem.* **269,** 8459–8465.

Fontes, E. P. B., Gladfelter, H. J., Schaffer, R. L., Petty, I. T. D., and Hanley-Bowdoin, L. (1994b). Geminivirus replication origins have a modular organization. *Plant Cell* **6,** 405–416.

Fontes, E. P. B., Luckow, V. A., and Hanley-Bowdoin, L. (1992). A geminivirus replication protein is a sequence specific DNA binding protein. *Plant Cell* **4,** 597–608.

Frischmuth, T., Roberts, S., Von Arnim, A., and Stanley, J. (1993). Specificity of bipartite geminivirus movement proteins. *Virology* **196,** 666–673.

Gladfelter, H. J., Eagle, P. A., Fontes, E. P. B., Batts, L., and Hanley-Bowdoin, L. (1997). Two domains of the AL1 protein mediate geminivirus origin recognition. *Virology* **239,** 186–197.

Heyraud-Nitschke, F., Schumacher, S., Laufs, J., Schaefer, S., Schell, J., and Gronenborn, B. (1995). Determination of the origin cleavage and joining domain of geminivirus Rep proteins. *Nucleic Acids Res.* **23,** 910–916.

Horton, R. M. (1994). *In vitro* recombination and mutagenesis of DNA. *In* "Methods in Molecular Biology: PCR Cloning Protocols" (B. A. White, Ed.), Vol. 67, pp.141–150. Humana Press, Totowa, NJ.

Jupin, I., Hericourt, F., Benz. B., and Gronenborn, B. (1995). DNA replication specificity of TYLCV geminivirus is mediated by the amino terminal 116 amino acids of the rep protein. *FEBS Lett.* **262,** 116–120.

Laufs, J., Schumacher, S., Geisler, N., Jupin, I., and Gronenborn, B. (1995a). Identification of the nicking tyrosine of the geminivirus Rep protein. *FEBS Lett.* **377,** 258–262.

Laufs, J., Traut, W., Heyraud, F., Matzeit, V., Rogers, S. G., Schell, J., and Gronenborn, B. (1995b). *In vitro* cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proc. Natl. Acad. Sci. USA* **92,** 3879–3883.

Lazarowitz, S. (1992). Geminiviruses: Genome structure and gene function. *Crit. Rev. Plant. Sci.* **11,** 329–349.

Lazarowitz, S. G., Wu, L. C., Rogers S. G., and Elmer, J. S. (1992). Sequence-specific interaction with the viral AL1 protein identifies a geminivirus DNA replication origin. *Plant Cell* **4,** 799–809.

Mettler, I. J. (1987). A simple and rapid method for minipreparation of DNA from tissue cultured plant cells. *Plant Mol. Biol. Rep.* **5,** 346–349.

Orozco, B. M., Gladfelter, H. J., Settlage, S. B., Eagle, P. A., Gentry, R. N., and Hanley-Bowdoin, L. (1998). Multiple cis elements contribute to geminivirus origin function. *Virology* **242,** 346–356.

Orozco, B. M., Miller, A. B., Settlage, S. B., and Hanley-Bowdoin, L. (1997). Functional domains of a geminivirus replication protein. *J. Biol. Chem.* **272,** 9840–9846.

Padidam, M., Beachy, R. N., and Fauquet, C. M. (1995). Tomato leaf curl geminivirus from India has a bipartite genome and coat protein is not essential for infectivity. *J. Gen. Virol.* **76,** 25–35.

Saunders, K., Lucy, A., and Stanley, J. (1991). DNA forms of the geminivirus African cassava mosaic consistent with a rolling circle mechanism of replication. *Nucleic Acids Res.* **19,** 2325–2330.

Smith, G. E., and Summers, M. D. (1978). Analysis of baculovirus genomes with restriction endonucleases. *Virology* **89,** 517–527.

Stanley, J. (1991). The molecular determinants of geminivirus pathogenesis. *Semin. Virol.* **2,** 139–150.

Stanley, J. (1995). Analysis of African cassava mosaic virus recombinants suggests that strand nicking occurs within the conserved nonanucleotide motif during the initiation of rolling circle replication. *Virology* **206,** 707–712.

Stanley, J., Frischmuth, T., and Ellwood, S. (1990). Defective viral DNA ameliorates symptoms of geminivirus infection in transgenic plants. *Proc. Natl. Acad. Sci. USA* **87,** 6291–6295.

Stenger, D. C. (1998). Replication specificity elements of the Worland strain of beet curly top virus are compatible with those of the CFH strain but not those of the Cal/Logan strain. *Phytopathology* **88,** 1174–1178.

Stenger, D. C., Revington, G. N., Stevenson, M. C., and Bisaro, D. M. (1991). Replicational release of geminiviral genomes from tandemly repeated copies: Evidence for rolling circle replication of a plant viral DNA. *Proc. Natl. Acad. Sci. USA* **88,** 8029–8033.

Summers, M. D., and Smith, G. E. (1987). A manual of methods for baculovirus vectors and insect cell cultures procedures. Texas Agricultural Experiment Station Bulletin No. 1555.

Sunter, G., Hartitz, M. D., and Bisaro, D. M. (1993). Tomato golden mosaic virus leftward gene expression: Autoregulation of geminivirus replication protein. *Virology* **179,** 69–77.

Towbin, H., Kemp, D. J., and Stark, G. R. (1979). Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose sheets: Procedures and some applications. *Proc. Natl. Acad. Sci. USA* **76,** 4350–4354.

Watanabe, Y., Meshi, T., and Okada, Y. (1987). Infection of tobacco protoplasts with *in vitro* transcribed tobacco mosaic virus RNA using an improved electroporation method. *FEBS Lett.* **219,** 65–69.